

CHINESE WORD SEGMENTATION

ABSTRACT OF THE DISCLOSURE

The present invention relates to a corpus
for use in training a language model. The corpus
5 includes a plurality of characters and a plurality of
morphological tags associated with a plurality of
sequences of characters. The plurality of
morphological tags indicate a morphological type of
an associated sequence of characters and a
10 combination of parts forming a morphological subtype.